



Bioinformatics

Week 5, 2020-04-01

Xiaofan Zhou

BLAST

- **Basic Local Alignment Searching Tool**
 - “(it) finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance”.

Why BLAST?

- To find homologs
- To find regions that are conserved (or not)
- To infer structure or function
-

BLAST

- **Basic Local Alignment Searching Tool**

- *“(it) finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance”.*

- **Pairwise sequence alignment**

- *“the process of lining up two sequences to achieve maximal levels of identity (and conservation, in the case of amino acid sequences)”.*

Target Sequence:

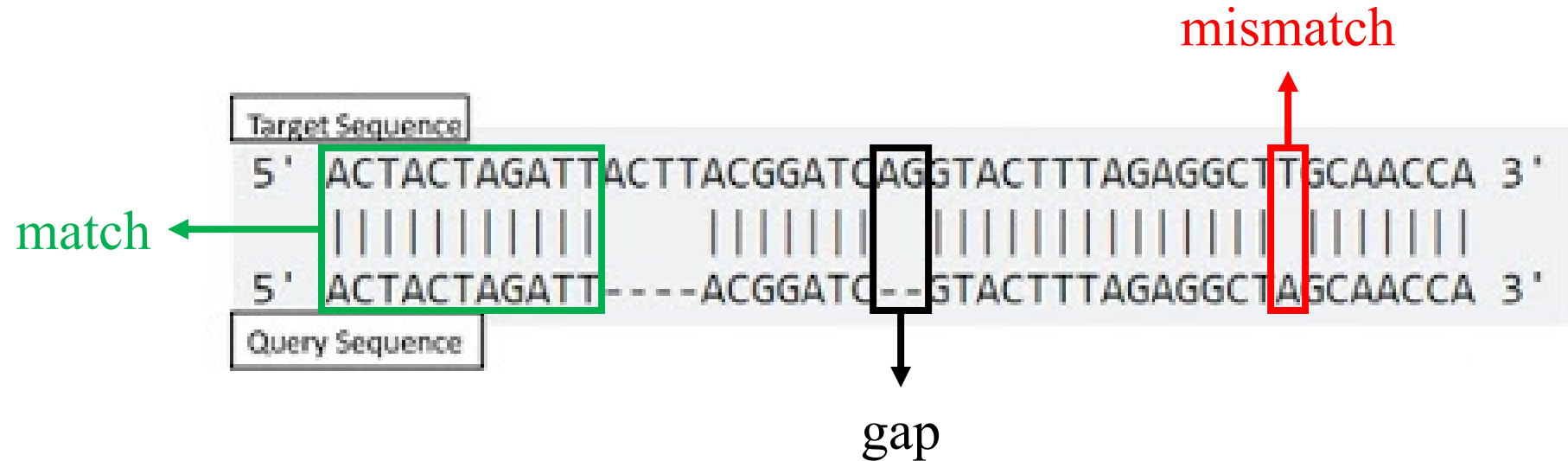
ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA

Query Sequence:

ACTACTAGATTACGGATCGTACTTTAGAGGCTAGCAACCA



Scoring of pairwise alignment: DNA



Default Mega-BLAST scoring matrix:

- match: +1
- mismatch: -2
- gap opening: 0
- gap extension: -2.5

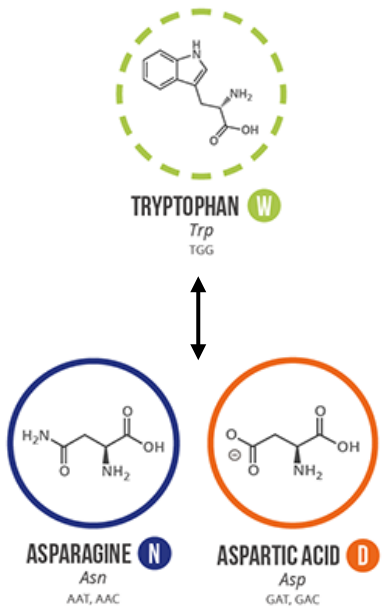
The above alignment has 39 matches,
1 mismatch, 6 gaps (2 openings):
 $= 39 * 1 + 1 * (-2) + 6 * (-2.5) + 2 * 0$
 $= 22$

Protein scoring matrix

Chart Key: ● ALIPHATIC ● AROMATIC ● ACIDIC ● BASIC ● HYDROXYLIC ● SULFUR-CONTAINING ● AMIDIC ○ NON-ESSENTIAL ○ ESSENTIAL

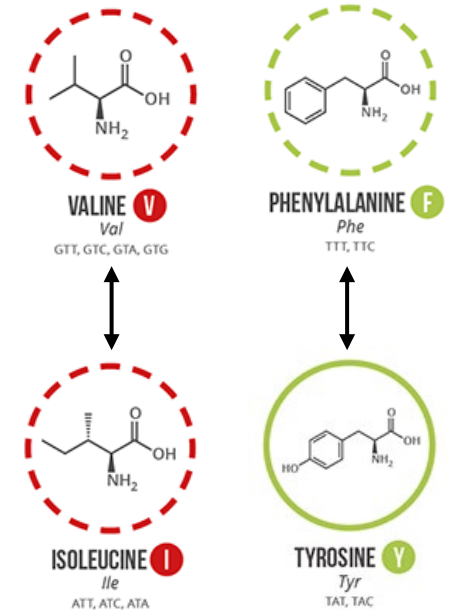
<p>Chemical Structure single letter code</p> <p>NAME A three letter code DNA codons</p>	<p>ALANINE A <i>Ala</i> GCT, GCC, GCA, GCG</p>	<p>GLYCINE G <i>Gly</i> GGT, GGC, GGA, GGG</p>	<p>ISOLEUCINE I <i>Ile</i> ATT, ATC, ATA</p>	<p>LEUCINE L <i>Leu</i> CTT, CTC, CTA, CTG, TTA, TTG</p>	<p>PROLINE P <i>Pro</i> CCT, CCC, CCA, CCG</p>	<p>VALINE V <i>Val</i> GTT, GTC, GTA, GTG</p>
<p>PHENYLALANINE F <i>Phe</i> TTT, TTC</p>	<p>TRYPTOPHAN W <i>Trp</i> TGG</p>	<p>TYROSINE Y <i>Tyr</i> TAT, TAC</p>	<p>ASPARTIC ACID D <i>Asp</i> GAT, GAC</p>	<p>GLUTAMIC ACID E <i>Glu</i> GAA, GAG</p>	<p>ARGININE R <i>Arg</i> CGT, CGC, CGA, CGG, AGA, AGG</p>	<p>HISTIDINE H <i>His</i> CAT, CAC</p>
<p>LYSINE K <i>Lys</i> AAA, AAG</p>	<p>SERINE S <i>Ser</i> TCT, TCC, TCA, TCG, AGT, AGC</p>	<p>THREONINE T <i>Thr</i> ACT, ACC, ACA, ACG</p>	<p>CYSTEINE C <i>Cys</i> TGT, TGC</p>	<p>METHIONINE M <i>Met</i> ATG</p>	<p>ASPARAGINE N <i>Asn</i> AAT, AAC</p>	<p>GLUTAMINE Q <i>Gln</i> CAA, CAG</p>

Protein scoring matrix



A	4																			
R	-1	5																		
N	-2	0	6																	
D	-2	-2	1	6																
C	0	-3	-3	-3	9															
Q	-1	1	0	0	-3	5														
E	-1	0	0	2	-4	2	5													
G	0	-2	0	-1	-3	-2	-2	6												
H	-2	0	1	-1	-3	0	0	-2	8											
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
K	-1	2	0	-1	-1	1	1	-2	-1	-3	-2	5								
M	-1	-2	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4
	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V

BLOSUM62, the default scoring matrix of BLASTP



Scoring of pairwise alignment: protein

Score = 18.1 bits (35), Expect = 0.015, Method: Composition-based stats.
Identities = 11/24 (45%), Positives = 12/24 (50%), Gaps = 2/24 (8%)

Query	12	VTALWGKVNVD--EVGGEALGRLL	33
		V +WGKV D G E L RL	
Sbjct	11	VLNVWGKVEADIPGHGQEV LIRLF	34

match	4	11	5	6	6	5	4	5	sum of matches: +60
			6	4				4	

mismatch	-1	1	0	-2	-2	-4	0	sum of mismatches: -13
	-2		0	-3	0			

gap open				-11				sum of gap penalties: -13
----------	--	--	--	-----	--	--	--	---------------------------

gap extend				-2				
------------	--	--	--	----	--	--	--	--

total raw score: 61 - 13 - 13 = 35

Global vs. local alignment

(a)

NP_824492.1	1	MCGDMTVHTVEYIRYRIPEQQSAEFLAAYTRAAQQLAAAPQCVDYELARC	50
NP_337032.1	1		0
NP_824492.1	51	EEDFEHFVLRLITWTSTEDHIEGFRKSELPDFLAEIRPYISSIEEMRHYK	100
NP_337032.1	1		0
NP_824492.1	101	PTTVRGTAAPVPTLYAWAGGAEAFARLTEVFYKVLKDDVLAPVFEGMAP	150
NP_337032.1	1	MEGMDQMPKSFYDAVGGAKTFDAIVSRFYAQVAEDEVLRVY----P	43
NP_824492.1	151	EH----AAHVALWLGEVFGGPAAYSETQGGHGHMVAHLGKNITEVQRR	195
NP_337032.1	44	EDDLAGAEERLRMFLEQYWGGPRTYSE-QRGHPRLMRHAPFRISLIERD	92
NP_824492.1	196	RWVNLLQDAADDAGLPT-DAEFRSAFLAYAEWGTRLAVYFSGPDVPPAE	244
NP_337032.1	93	AWLRMHTAVASIDSETLDDEHRRELLDYLEMAAHSLV--NSPF	134
NP_824492.1	245	QPVPQWSWGAMPPYQP	260
NP_337032.1	135		134

global alignment – 15% identity

(b)

NP_824492.1	113	TLYAWAGGAEAFARLTEVFYKVLKDDVLAPVFEGMAPEH-----AAHVA	157
NP_337032.1	10	SFYDAVGGAKTFDAIVSRFYAQVAEDEVLRVY----PEDDLAGAEERLR	55
NP_824492.1	158	LWLGEVFGGPAAYSETQGGHGHMVAHLGKNITEVQRRRWVNLLQDAADD	207
NP_337032.1	56	MFLEQYWGGPRTYSE-QRGHPRLMRHAPFRISLIERDAWLRCMHTAVAS	104
NP_824492.1	208	AGLPT-DAEFRSAFLAYAE	225
NP_337032.1	105	IDSETLDDEHRRELLDYLE	123

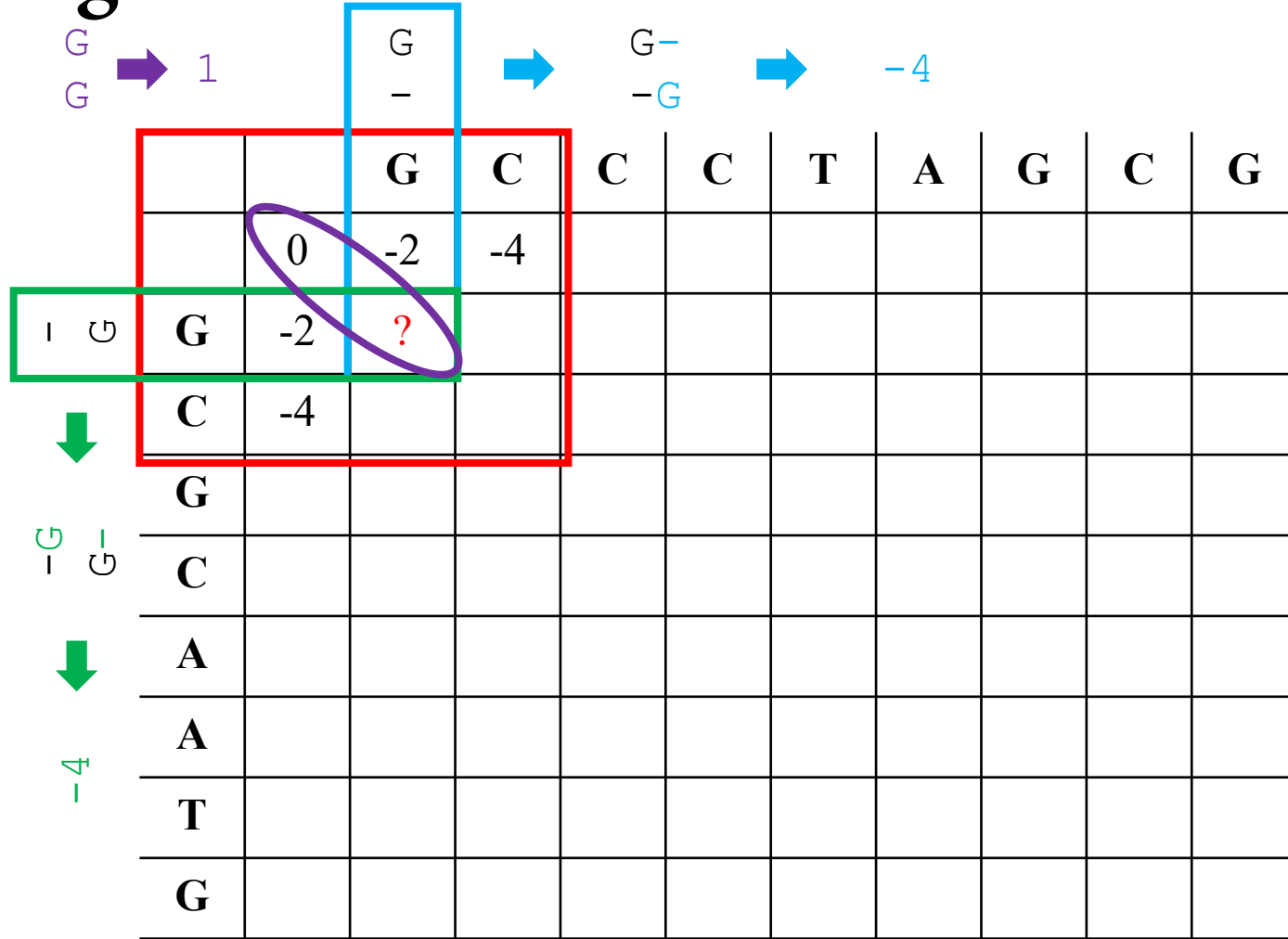
local alignment – 30% identity

Global alignment: Needleman-Wunsch

		G	C									
		-	-									
I	U			G	C	C	C	T	A	G	C	G
			0	-2	-4							
I	U	G	-2	?								
	U	C	-4									
		G										
		C										
		A										
		A										
		T										
		G										

- match: +1
- mismatch: -1
- gap opening: 0
- gap extension: -2

Global alignment: Needleman-Wunsch



Global alignment: Needleman-Wunsch

[illegible]

Global alignment: Needleman-Wunsch

		G	C	C	C	T	A	G	C	G
	0	-2	-4	-6	-8	-10	-12	-14	-16	-18
G	-2	1	-1	-3	-5	-7	-9	-11	-13	-15
C	-4	-1	2	0	-2	-4	-6	-8	-10	-12
G	-6	-3	0	1	-1	-3	-5	-5	-7	-9
C	-8	-5	-2	1	2	0	-2	-4	-4	-6
A	-10	-7	-4	-1	0	1	1	-1	-3	-5
A	-12	-9	-6	-3	-2	-1	2	0	-2	-4
T	-14	-11	-8	-5	-4	-1	0	1	-1	-3
G	-16	-13	-10	-7	-6	-3	-2	1	0	0

Global alignment: Needleman-Wunsch

		G	C	C	C	T	A	G	C	G
	0	-2	-4	-6	-8	-10	-12	-14	-16	-18
G	-2	1	-1	-3	-5	-7	-9	-11	-13	-15
C	-4	-1	2	0	-2	-4	-6	-8	-10	-12
G	-6	-3	0	1	-1	-3	-5	-5	-7	-9
C	-8	-5	-2	1	2	0	-2	-4	-4	-6
A	-10	-7	-4	-1	0	1	1	-1	-3	-5
A	-12	-9	-6	-3	-2	-1	2	0	-2	-4
T	-14	-11	-8	-5	-4	-1	0	1	-1	-3
G	-16	-13	-10	-7	-6	-3	-2	1	0	0

Best global alignment:

GCCCTAGCG
||| |
GCGC-AATG

alignment core: 0

Local alignment: Smith-Waterman

		G	C	C	C	T	A	G	C	G
	0	0	0	0	0	0	0	0	0	0
G	0	1	0	0	0	0	0	1	0	1
C	0	0	2	1	1	0	0	0	2	0
G	0	1	0	1	0	0	0	1	0	3
C	0	0	2	1	2	0	0	0	2	1
A	0	0	0	1	0	1	1	0	0	1
A	0	0	0	0	0	0	2	0	0	0
T	0	0	0	0	0	1	0	1	0	0
G	0	1	0	0	0	0	0	1	0	1

Local alignment: Smith-Waterman

		G	C	C	C	T	A	G	C	G
	0	0	0	0	0	0	0	0	0	0
G	0	1	0	0	0	0	0	1	0	1
C	0	0	2	1	1	0	0	0	2	0
G	0	1	0	1	0	0	0	1	0	3
C	0	0	2	1	2	0	0	0	2	1
A	0	0	0	1	0	1	1	0	0	1
A	0	0	0	0	0	0	2	0	0	0
T	0	0	0	0	0	1	0	1	0	0
G	0	1	0	0	0	0	0	1	0	1

Best local alignment:

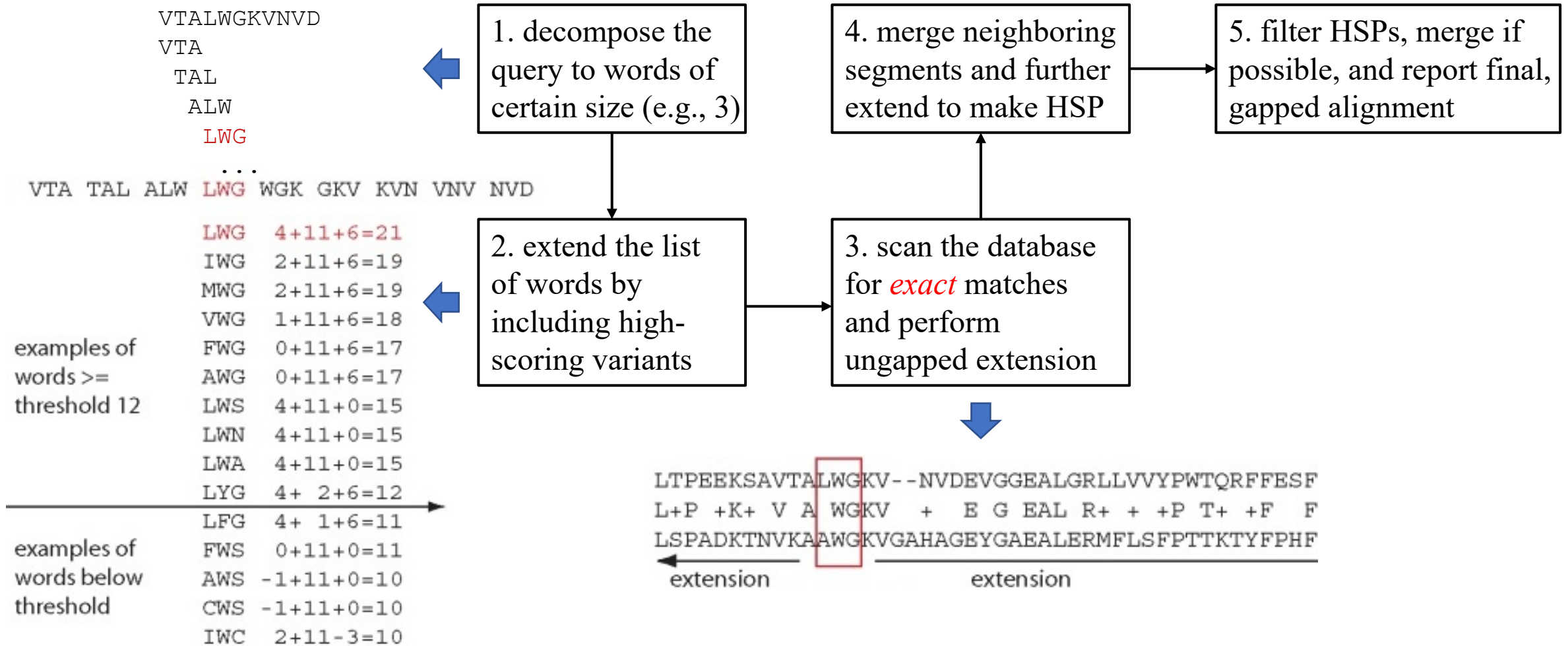
GCCCTAGCG
 | | |
 GCGCAATG

alignment core: 3

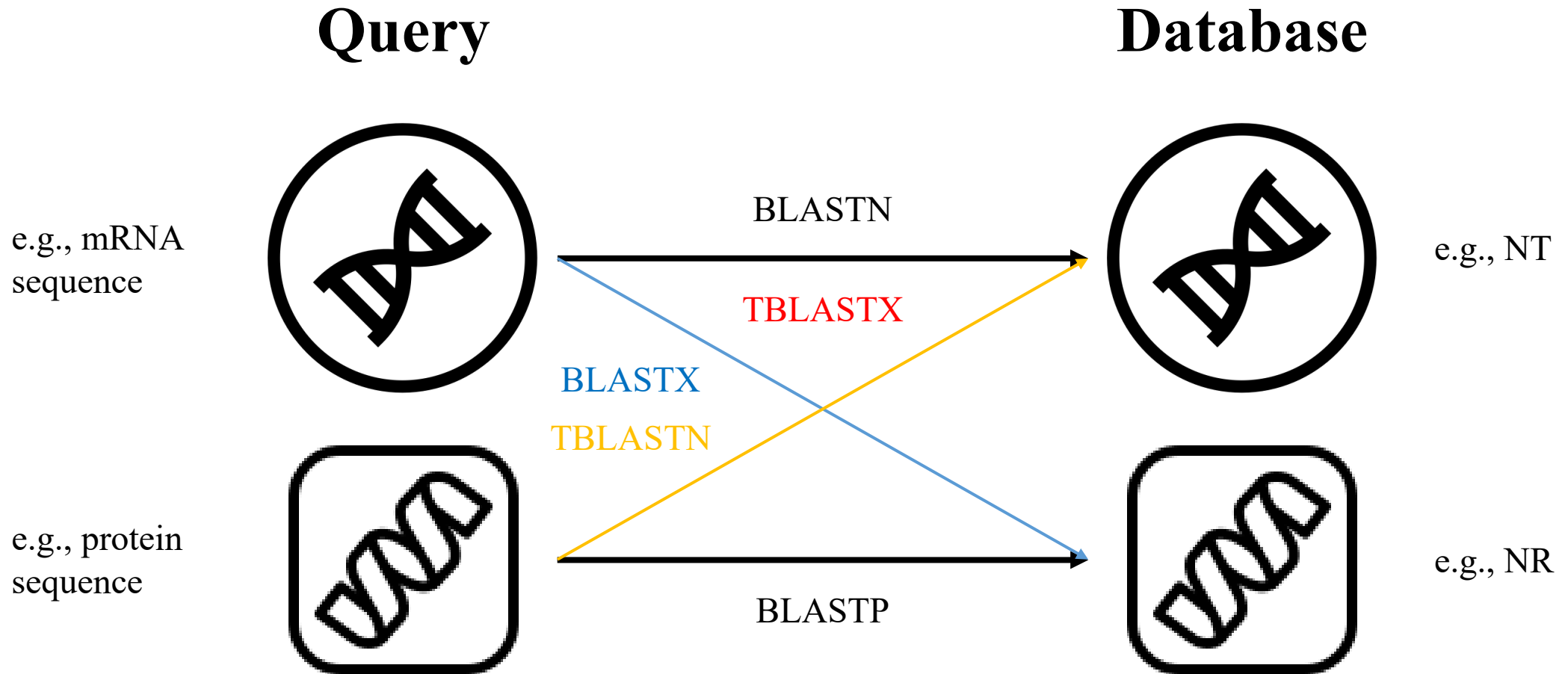
BLAST

- **Basic Local Alignment Searching Tool**
 - *“(it) finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance”.*
- One of the most highly cited bioinformatics programs
 - original BLAST (1990): >83,000 citations
 - gapped BLAST and PSI-BLAST (1997): >73,000 citations
 - BLAST+ (2008): ~7,000 citations
- Why is BLAST so popular?

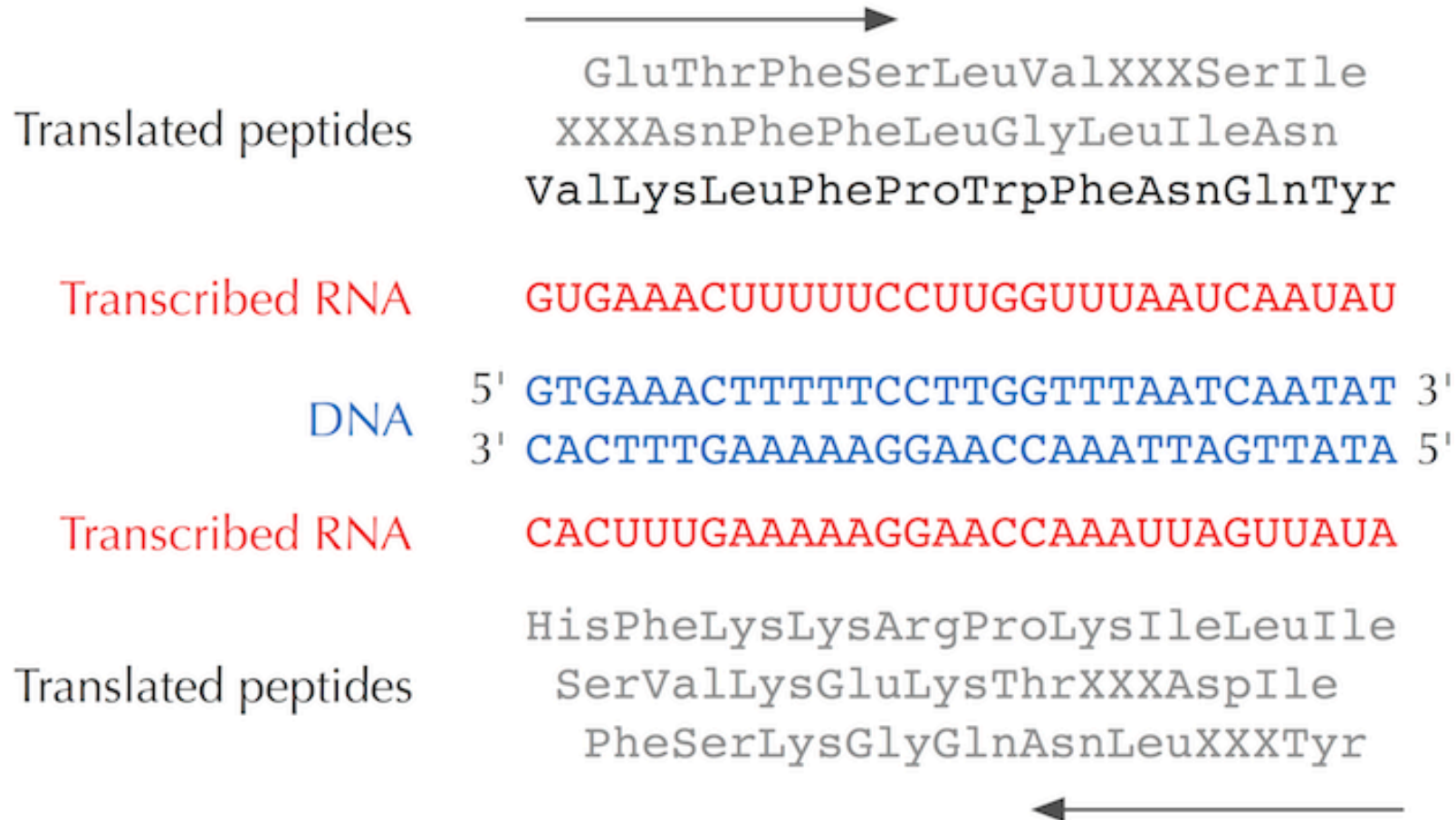
How does BLAST work?



BLAST programs

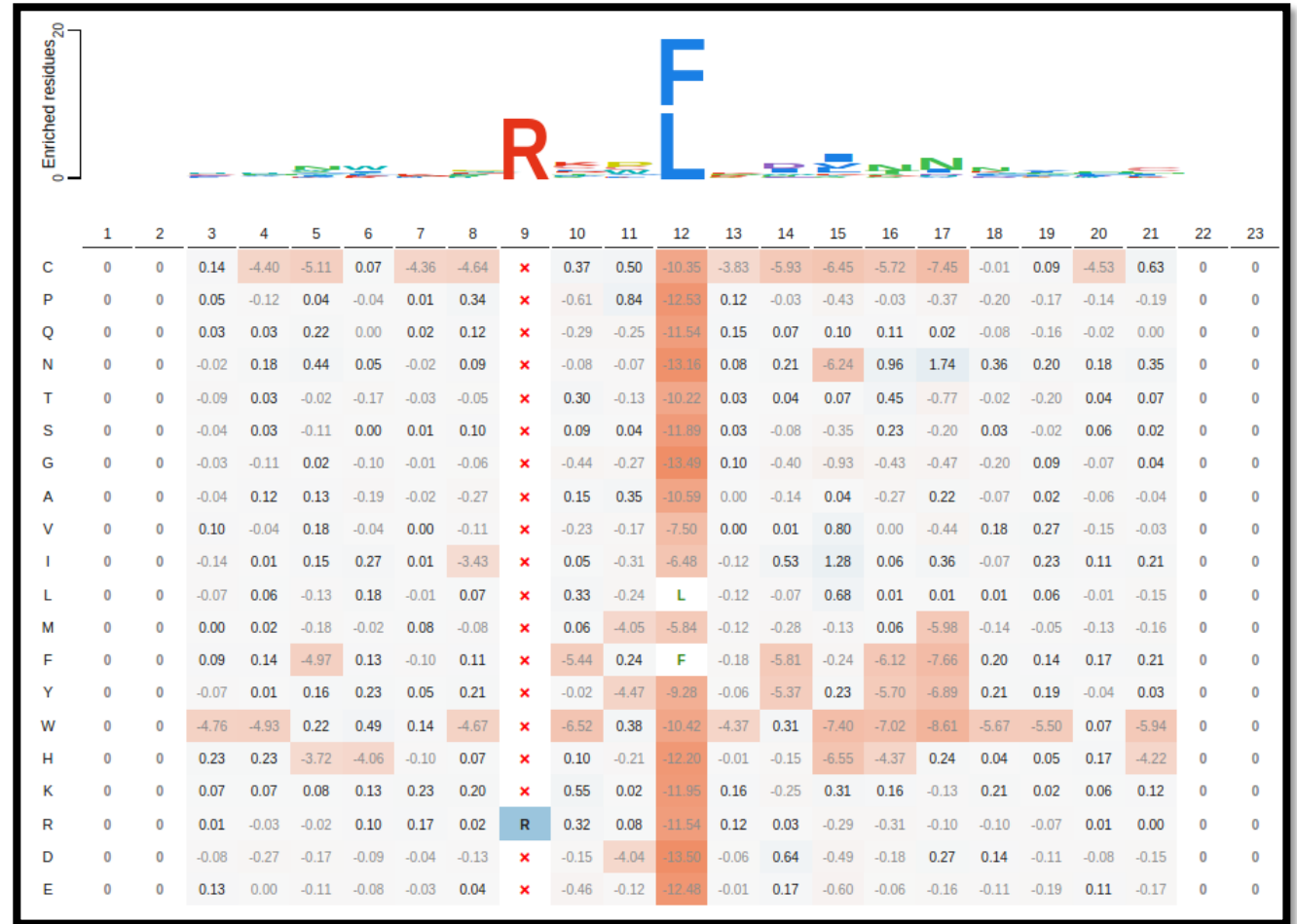


Six-frame translation



PSI-BLAST and DELTA-BLAST

- sites in a protein are not all the same
- Positional Specific Scoring Matrix (PSSM)
- achieves higher sensitivity



Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.

[Learn more](#)

NEWS

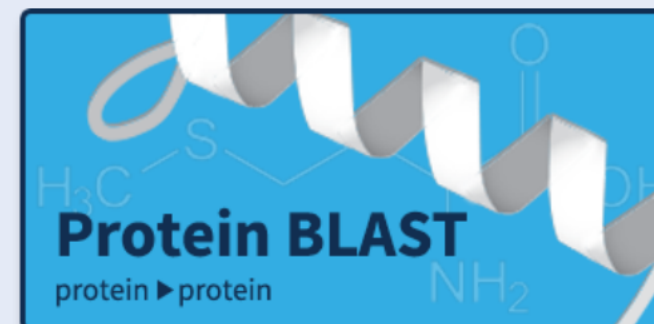
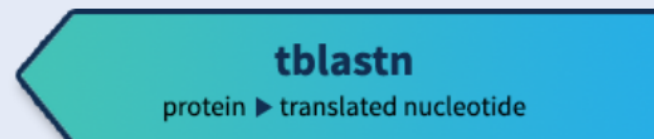
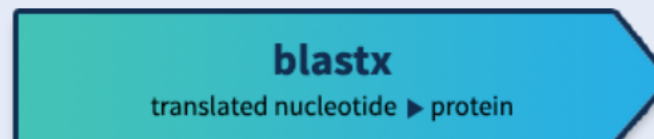
A New version of Magic-BLAST(1.5.0) is here.

The BLAST tool for mapping large next-generation RNA or DNA sequencing runs against a whole genome or transcriptome.

Wed, 28 Aug 2019 17:00:00 EST

[More BLAST news...](#)

Web BLAST



BLAST Genomes

[Search](#)[Human](#)[Mouse](#)[Rat](#)[Microbes](#)

BLAST®

>> blastp suite

Standard Protein BLAST

blastn

blastp

blastx

tblastn

tblastx

BLASTP programs search protein databases using a protein query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) ?

>NP_001243779.1 titin isoform N2BA [Homo sapiens]
MTTQAPTFTQPLQSVVLEGSTATFEAHISGFFVPEVSWFRDQVISTSTLPGVQISFSDGRAKLT
PAV
TKANSGRYSKATNGSGQATSTAELLVKAETAPPNFVQRLQSMQVTRQGSQVRLQVRVTGIPVVKF
YRD

Clear

Query subrange ?

From

- Non-redundant protein sequences (nr)
- Non-redundant protein sequences (nr)
- Reference proteins (refseq_protein)
- Model Organisms (landmark)
- UniProtKB/Swiss-Prot (swissprot)
- Patented protein sequences (pataa)
- Protein Data Bank proteins (pdb)
- Metagenomic proteins (env_nr)
- Transcriptome Shotgun Assembly proteins (tsa_nr)

Or, upload file

选择文件

未选择任何文件

Job Title

NP_001243779.1 titin isoform N2BA [Homo sapiens]

Enter a descriptive title for your BLAST search ?

☐ Align two or more sequences ?

Choose Search Set

Database

Non-redundant protein sequences (nr)

Organism

Optional

mus

☐ exclude

+

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. ?

Exclude

Optional

☐ Models (XM/XP)

☐ Non-redundant RefSeq proteins (WP)

☐ Uncultured/environmental samples


Program Selection

- Algorithm
- ☐ Quick BLASTP (Accelerated protein-protein BLAST)
- ☒ blastp (protein-protein BLAST)
- ☐ PSI-BLAST (Position-Specific Iterated BLAST)
- ☐ PHI-BLAST (Pattern Hit Initiated BLAST)
- ☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)
- Choose a BLAST algorithm ?

- Nucleotide collection (nr/nt)
- Genomic plus Transcript
- Human genomic plus transcript (Human G+T)
- Mouse genomic plus transcript (Mouse G+T)
- Other Databases
- Nucleotide collection (nr/nt)
- 16S ribosomal RNA sequences (Bacteria and Archaea)
- Reference RNA sequences (refseq_rna)
- RefSeq Representative genomes (refseq_representative_genomes)
- RefSeq Genome Database (refseq_genomes)
- Whole-genome shotgun contigs (wgs)
- Expressed sequence tags (est)
- Sequence Read Archive (SRA)
- Transcriptome Shotgun Assembly (TSA)
- High throughput genomic sequences (HTGS)
- Patent sequences (pat)
- Protein Data Bank (pdb)
- Reference genomic sequences (refseq_genomic)
- Human RefSeqGene sequences (RefSeq_Gene)
- Genomic survey sequences (gss)
- Sequence tagged sites (dbsts)

- mus
- Mus (taxid:10088)
- Mus (taxid:10088)
- Mus (taxid:862507)
- Mus muscaris (taxid:10090)
- Mus musculus (taxid:10090)
- Mus sp. 129SV (taxid:10090)
- Mus laniger (taxid:34839)
- Mus laniger Molina, 1782 (taxid:34839)
- Mus caroli (taxid:10089)
- Mus formosanus (taxid:10089)
- Mus pahari (taxid:10093)
- Mus domesticus (taxid:10092)

☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm 

BLAST

Search **database nr** using **Blastp (protein-protein BLAST)**

☒ Show results in a new window


 [Algorithm parameters](#)

[Restore default search parameters](#)

General Parameters

Max target sequences

100 ▾

Select the maximum number of aligned sequences to display 

Short queries

☒ Automatically adjust parameters for short input sequences 

Expect threshold

10



Word size

6 ▾



Max matches in a query range

0



Scoring Parameters

Matrix

BLOSUM62 ▾



Gap Costs

Existence: 11 Extension: 1 ▾



Compositional adjustments

Conditional compositional score matrix adjustment ▾





Filters and Masking

Filter

☐ Low complexity regions 

Mask

☐ Mask for lookup table only 

☐ Mask lower case letters 

BLAST

Search **database nr** using **Blastp (protein-protein BLAST)**

☒ Show results in a new window



Format Request Status

[\[Formatting options\]](#)

Job Title: NP_001243779.1 titin isoform N2BA [Homo sapiens]

Request ID	RGTXK3RT015
Status	Searching
Submitted at	Tue Sep 10 23:19:15 2019
Current time	Tue Sep 10 23:19:26 2019
Time since submission	00:00:10

This page will be automatically updated in 2 seconds

BLAST is a registered trademark of the National Library of Medicine

[Support center](#) [Mailing list](#) 

NCBI
National Center for Biotechnology Information, U.S. National Library of Medicine
8600 Rockville Pike, Bethesda MD, 20894 USA

[Policies and Guidelines](#) | [Contact](#)





[← Edit Search](#)

[Save Search](#)

[Search Summary](#) ▼

[? How to read this report?](#)

[▶ BLAST Help Videos](#)

[↶ Back to Traditional Results Page](#)

i Your search is limited to records that include: Mus (taxid:10088)

Job Title NP_001243779.1 titin isoform N2BA [Homo sapiens]
RID RGTK3RT015 [Search expires on 09-12 11:19 am](#) [Download All](#) ▼
Program BLASTP [?](#) [Citation](#) ▼
Database nr [See details](#) ▼
Query ID lcl|Query_260495
Description NP_001243779.1 titin isoform N2BA [Homo sapiens]
Molecule type amino acid
Query Length 34350
Other reports [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#) [?](#)

Filter Results

Organism only top 20 will appear

☐ exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity

to

E value

to

Filter

Reset

Descriptions

[Graphic Summary](#)

[Alignments](#)

[Taxonomy](#)

Sequences producing significant alignments

[Download](#) ▼

[Manage Columns](#) ▼

Show

100 ▼



☒ select all 100 sequences selected

[GenPept](#)

[Graphics](#)

[Distance tree of results](#)

[Multiple alignment](#)

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	titin isoform N2-A [Mus musculus]	44050	1.270e+05	96%	0.0	93.16%	NP_035782.3
<input checked="" type="checkbox"/>	titin isoform X47 [Mus caroli]	43983	67215	93%	0.0	93.03%	XP_029329032.1
<input checked="" type="checkbox"/>	titin isoform X3 [Mus pahari]	43769	1.126e+05	100%	0.0	93.90%	XP_029391868.1
<input checked="" type="checkbox"/>	titin isoform X1 [Mus musculus]	43745	1.260e+05	100%	0.0	92.66%	XP_017172628.1
<input checked="" type="checkbox"/>	RecName: Full=Titin; AltName: Full=Connectin [Mus musculus]	43735	1.163e+05	100%	0.0	92.86%	A2ASS6.1
<input checked="" type="checkbox"/>	titin isoform X37 [Mus caroli]	43699	89404	98%	0.0	92.43%	XP_029328975.1
<input checked="" type="checkbox"/>	titin isoform X42 [Mus caroli]	43697	1.093e+05	98%	0.0	92.43%	XP_029329006.1
<input checked="" type="checkbox"/>	titin isoform X39 [Mus caroli]	43697	1.116e+05	98%	0.0	92.43%	XP_029328983.1
<input checked="" type="checkbox"/>	titin isoform X40 [Mus caroli]	43697	1.113e+05	98%	0.0	92.43%	XP_029328994.1
<input checked="" type="checkbox"/>	titin [Mus musculus]	43693	80634	99%	0.0	93.54%	EDL27217.1
<input checked="" type="checkbox"/>	titin isoform X4 [Mus caroli]	43690	89120	98%	0.0	92.55%	XP_029328892.1
<input checked="" type="checkbox"/>	titin isoform X35 [Mus caroli]	43690	1.158e+05	98%	0.0	92.43%	XP_029328964.1



[← Edit Search](#)

[Save Search](#)

[Search Summary](#) ▼

[? How to read this report?](#)

[▶ BLAST Help Videos](#)

[↶ Back to Traditional Results Page](#)

i Your search is limited to records that include: Mus (taxid:10088)

Job Title	NP_001243779.1 titin isoform N2BA [Homo sapiens]
RID	RGTXK3RT015 <small>Search expires on 09-12 11:19 am</small> Download All ▼
Program	BLASTP ? Citation ▼
Database	nr See details ▼
Query ID	lcl Query_260495
Description	NP_001243779.1 titin isoform N2BA [Homo sapiens]
Molecule type	amino acid
Query Length	34350
Other reports	Distance tree of results Multiple alignment MSA viewer ?

Filter Results

Organism only top 20 will appear

☐ exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity

 to

E value

 to

Filter

Reset

[Descriptions](#)

[Graphic Summary](#)

[Alignments](#)

[Taxonomy](#)

[hover to see the title](#)

[click to show alignments](#)

☒ Show Conserved Domains

Alignment Scores

☐ < 40

☐ 40 - 50

☐ 50 - 80

☐ 80 - 200

☐ >= 200

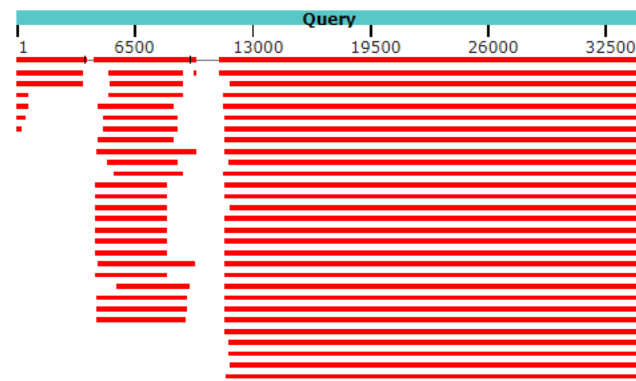
[?](#)

100 sequences selected [?](#)

Hit the button to

[See conserved domains from CDD](#)

Distribution of the top 200 Blast Hits on 100 subject sequences





< Edit Search

Save Search

Search Summary ▾

How to read this report?

BLAST Help Videos

Back to Traditional Results Page

i Your search is limited to records that include: Mus (taxid:10088)

Job Title	NP_001243779.1 titin isoform N2BA [Homo sapiens]		
RID	RGTXK3RT015	Search expires on 09-12 11:19 am	Download All ▾
Program	BLASTP	Citation ▾	
Database	nr	See details ▾	
Query ID	lcl Query_260495		
Description	NP_001243779.1 titin isoform N2BA [Homo sapiens]		
Molecule type	amino acid		
Query Length	34350		
Other reports	Distance tree of results	Multiple alignment	MSA viewer i

Filter Results

Organism *only top 20 will appear*

☐ exclude

Type common name, binomial, taxid or group name

+ [Add organism](#)

Percent Identity

E value

 to to

Filter

Reset

Descriptions

Graphic Summary

Alignments

Taxonomy

Alignment view Pairwise ▾



Download ▾

100 sequences selected **i**

Download ▾

[GenPept](#) [Graphics](#)

Sort by: E value ▾

Next Previous Descriptions

titin isoform N2-A [Mus musculus]

Sequence ID: [NP_035782.3](#) Length: 33467 Number of Matches: 813

Range 1: 10251 to 33467 [GenPept](#) [Graphics](#)

Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
44050 bits(114345)	0.0	Compositional matrix adjust.	21681/23273(93%)	22507/23273(96%)	112/23273(0%)
Query 11134	APPKAVPEVPKKVVEEKRIILPKKEEVLPVEV-----TEEPPEEPISSEEEIPEEPPSIE				
	APP KVPEV KKVEE+RII PKKEEV P EV T E E E E E				
Sbjct 10251	APPTKVPEVSKKVEERRIIPKKEEVPPAEVYEEAEPTPEEIPEEPPSIEEEEIVEEEE				
Query 11187	EVEEVAPPRVPEVIKKAVPEAPTVPKKVEAPPAKVSKKIPEEKVPVPVQKKEAPPAKVP				
	E EEV PPR PEV+KKAVPEAPTVPKK EAPPAKV KKIPEEKVPVPVQKKEAPPAKVP				
Sbjct 10311	EEEEVLPPRAPEVVKKAVPEAPTVPKKAEAPPAKVPKKIPEEKVPVPVQKKEAPPAKVP				
Query 11247	EVPKKVPEKK-----VLVPKKEAVPPAKGRTVLEEKVSVAFRQEVVVKERLELEVVEAE				
	EVPKKVPEKK V VPKKEAVPPAKG+ V EEK+SVA++QE +V+ER+ELE+VEA+				
Sbjct 10371	EVPKKVPEKKKIPEKKVPVPKKAEVPPAKGKAVFEKISVAYQQEELVQERIELELVEAK				

Related Information

[Gene](#) - associated gene details
[Genome Data Viewer](#) - aligned genomic context



[← Edit Search](#)

[Save Search](#)

[Search Summary](#) ▼

[? How to read this report?](#)

[▶ BLAST Help Videos](#)

[↶ Back to Traditional Results Page](#)

i Your search is limited to records that include: Mus (taxid:10088)

Job Title NP_001243779.1 titin isoform N2BA [Homo sapiens]
RID [RGTXK3RT015](#) *Search expires on 09-12 11:19 am* [Download All](#) ▼
Program BLASTP [?](#) [Citation](#) ▼
Database nr [See details](#) ▼
Query ID lcl|Query_260495
Description NP_001243779.1 titin isoform N2BA [Homo sapiens]
Molecule type amino acid
Query Length 34350
Other reports [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#) [?](#)

Filter Results

Organism *only top 20 will appear*

☐ exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity

to

E value

to

Filter

Reset

Descriptions

Graphic Summary

Alignments

Taxonomy

Reports

Lineage

Organism

Taxonomy

100 sequences selected [?](#)

Organism	Blast Name	Score	Number of Hits	Description
Mus	rodents		101	
. Mus	rodents		93	
. . Mus musculus	rodents	44050	34	Mus musculus hits
. . Mus caroli	rodents	43983	59	Mus caroli hits
. Mus pahari	rodents	43769	8	Mus pahari hits



Blog



Support Center

Critical BLAST parameters

- Scoring scheme:
 - scoring matrix
 - gap costs
- Word size
- *E*-value

How to interpret BLAST results

- Expect value: “*the number of high-scoring segment pairs (HSPs) expected to occur with a score of at least S .*”

Karlin Altschul statistics

$$E = -\ln(1-P)$$

$$E = K m n e^{-\lambda S} \longrightarrow \text{alignment score}$$

lengths of the query
and the database

Questions?